1

1

Introduction

Engineering disciplines are those fields of research and development that attempt to create products and systems operating in, and dealing with, the real world. The number of disciplines is large, as is the range of scales that they typically operate in: from the very small scale of nanotechnology up to very large scales that span whole regions, for example water management systems, electric power distribution systems or even global systems (e.g. the global positioning system, GPS). The level of advancement in the fields also varies wildly, from emerging techniques (again, nanotechnology) to trusted techniques that have been applied for centuries (architecture, hydraulic works). Nonetheless, the disciplines share one important aspect: engineering aims at designing and manufacturing systems that interface with the world around them.

Systems designed by engineers are often meant to influence their environment: to manipulate it, to move it, to stabilize it, to please it, and so on. To enable such actuation, these systems need information, for example values of physical quantities describing their environments and possibly also describing themselves. Two types of information sources are available: *prior* knowledge and *empirical* knowledge. The latter is knowledge obtained by sensorial observation. Prior knowledge is the knowledge that was already there before a given observation became available (this does not imply that prior knowledge is obtained without any observation). The combination of prior knowledge and empirical knowledge leads to *posterior* knowledge.

The sensory subsystem of a system produces measurement signals. These signals carry the empirical knowledge. Often, the direct usage of these signals is not possible, or is inefficient. This can have several causes:

- The information in the signals is not represented in an explicit way. It is often hidden and only available in an indirect, encoded, form.
- Measurement signals always come with noise and other hard-to-predict disturbances.
- The information brought forth by posterior knowledge is more accurate and more complete than information brought forth by empirical knowledge alone. Hence, measurement signals should be used in combination with prior knowledge.

Measurement signals need processing in order to suppress the noise and to disclose the information required for the task at hand.

1.1   The Scope of the Book

In a sense, classification and estimation deal with the same problem: given the measurement signals from the environment, how can the information that is needed for a system to operate in the real world be inferred? In other words, how should the measurements from a sensory system be processed in order to bring maximal information in an explicit and usable form? This is the main topic of this book.

Good processing of the measurement signals is possible only if some knowledge and understanding of the environment and the sensory system is present. Modelling certain aspects of that environment – like objects, physical processes or events – is a necessary task for the engineer. However, straightforward modelling is not always possible. Although the physical sciences provide ever deeper insight into nature, some systems are still only partially understood; just think of the weather. Even if systems are well understood, modelling them exhaustively may be beyond our current capabilities (i.e. computer power) or beyond the scope of the application. In such cases, approximate general models,

Figure 1.1 Licence plate recognition: a classification problem with noisy measurements.

but adapted to the system at hand, can be applied. The development of such models is also a topic of this book.

1.1.1 Classification

The title of the book already indicates the three main subtopics it will cover: classification, parameter estimation and state estimation. In classification, one tries to assign a class label to an object, a physical process or an event. Figure 1.1 illustrates the concept. In a speeding detector, the sensors are a radar speed detector and a high-resolution camera, placed in a box beside a road. When the radar detects a car approaching at too high a velocity (a parameter estimation problem), the camera is signalled to acquire an image of the car. The system should then recognize the licence plate, so that the driver of the car can be fined for the speeding violation. The system should be robust to differences in car model, illumination, weather circumstances, etc., so some pre-processing is necessary: locating the licence plate in the image, segmenting the individual characters and converting it into a binary image. The problem then breaks down to a number of individual classification problems. For each of the locations on the license plate, the input consists of a binary image of a character, normalized for size, skew/rotation and intensity. The desired output is the label of the true character, that is one of 'A', 'B',…, 'Z', '0',…, '9'.

Detection is a special case of classification. Here, only two class labels are available, for example 'yes' and 'no'. An example is a quality control system that approves the products of a manufacturer or refuses them. A second problem closely related to classification is identification: the act of proving that an object-under-test and a second object that is previously seen are the same. Usually, there is a large database of previously seen objects to choose from. An example is biometric identification, for example fingerprint recognition or face recognition. A third problem that can be solved by classification-like techniques is retrieval from a database, for example finding an image in an image database by specifying image features.

1.1.2 Parameter Estimation

In parameter estimation, one tries to derive a parametric description for an object, a physical process or an event. For example, in a beacon-based position measurement system (Figure 1.2), the goal is to find the position of an object, for example a ship or a mobile robot. In the two-dimensional case, two beacons with known reference positions suffice. The sensory system provides two measurements: the distances from the beacons to the object, r_1 and r_2. Since the position of the object involves two parameters, the estimation seems to boil down to solving two equations with two unknowns. However,
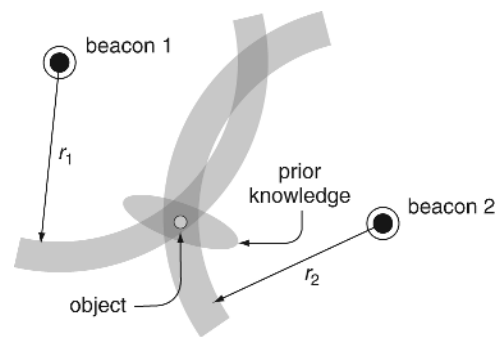

Figure 1.2 Position measurement: a parameter estimation problem handling uncertainties.

the situation is more complex because measurements always come with uncertainties. Usually, the application not only requires an estimate of the parameters but also an assessment of the uncertainty of that estimate. The situation is even more complicated because some prior knowledge about the position must be used to resolve the ambiguity of the solution. The prior knowledge can also be used to reduce the uncertainty of the final estimate.

In order to improve the accuracy of the estimate the engineer can increase the number of (independent) measurements to obtain an overdetermined system of equations. In order to reduce the cost of the sensory system, the engineer can also decrease the number of measurements, leaving us with fewer measurements than parameters. The system of equations is then underdetermined, but estimation is still possible if enough prior knowledge exists or if the parameters are related to each other (possibly in a statistical sense). In either case, the engineer is interested in the uncertainty of the estimate.

1.1.3   State Estimation

In state estimation, one tries to do either of the following – either assigning a class label or deriving a parametric (real-valued) description – but for processes that vary in time or space. There is a fundamental difference between the problems of classification and parameter estimation, on the one hand, and state estimation, on the other hand. This is the ordering in time (or space) in state estimation, which is absent from classification and parameter estimation. When no ordering in the data is assumed, the data can be processed in any order. In time series, ordering in time is essential for the process. This results in a fundamental difference in the treatment of the data.

In the discrete case, the states have discrete values (classes or labels) that are usually drawn from a finite set. An example of such a set is the alarm stages in a safety system (e.g. 'safe', 'pre-alarm', 'red alert', etc.). Other examples of discrete state estimation are speech recognition, printed or handwritten text recognition and the recognition of the operating modes of a machine.
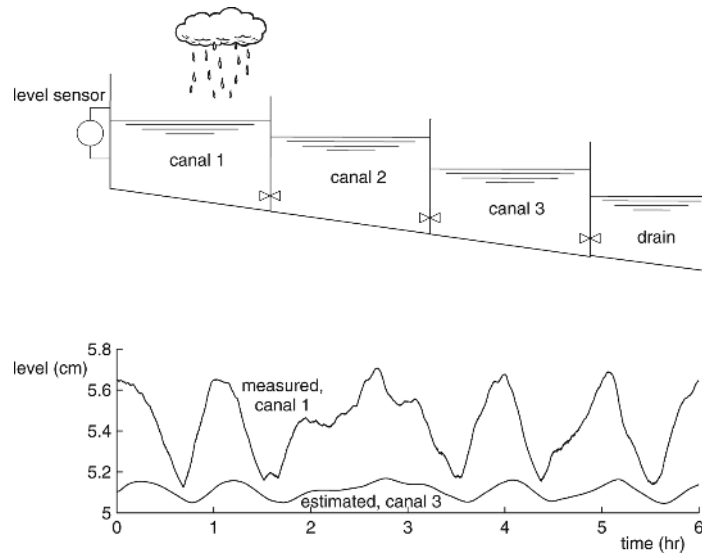
Figure 1.3 Assessment of water levels in a water management system: a state estimation problem (the data are obtained from a scale model).

An example of real-valued state estimation is the water management system of a region. Using a few level sensors and an adequate dynamical model of the water system, a state estimator is able to assess the water levels even at locations without level sensors. Short-term prediction of the levels is also possible. Figure 1.3 gives a view of a simple water management system of a single canal consisting of three linearly connected compartments. The compartments are filled by the precipitation in the surroundings of the canal. This occurs randomly but with a seasonal influence. The canal drains its water into a river. The measurement of the level in one compartment enables the estimation of the levels in all three compartments. For that, a dynamic model is used that describes the relations between flows and levels. Figure 1.3 shows an estimate of the level of the third compartment using measurements of the level in the first compartment. Prediction of the level in the third compartment is possible due to the causality of the process and the delay between the levels in the compartments.

1.1.4   Relations between the Subjects

The reader who is familiar with one or more of the three subjects might wonder why they are treated in one book. The three subjects share the following factors:

- In all cases, the engineer designs an instrument, that is a system whose task is to extract information about a real-world object, a physical process or an event.
- For that purpose, the instrument will be provided with a sensory subsystem that produces measurement signals. In all cases, these signals are represented by vectors (with fixed dimension) or sequences of vectors.
- The measurement vectors must be processed to reveal the information that is required for the task at hand.
- All three subjects rely on the availability of models describing the object/physical process/event and of models describing the sensory system.
- Modelling is an important part of the design stage. The suitability of the applied model is directly related to the performance of the resulting classifier/estimator.

Since the nature of the questions raised in the three subjects is similar, the analysis of all three cases can be done using the same framework. This allows an economical treatment of the subjects. The framework that will be used is a probabilistic one. In all three cases, the strategy will be to formulate the posterior knowledge in terms of a conditional probability (density) function:

$$P(\text{quantities of interest measurements available})$$

This so-called posterior probability combines the prior knowledge with the empirical knowledge by using Bayes' theorem for conditional probabilities. As discussed above, the framework is generic for all three cases. Of course, the elaboration of this principle for the three cases leads to different solutions because the nature of the 'quantities of interest' differs.

The second similarity between the topics is their reliance on models. It is assumed that the constitution of the object/physical process/event (including the sensory system) can be captured by a mathematical model. Unfortunately, the physical structures responsible for generating the objects/process/events are often

unknown, or at least partly unknown. Consequently, the model is also, at least partly, unknown. Sometimes, some functional form of the model is assumed, but the free parameters still have to be determined. In any case, empirical data are needed in order to establish the model, to tune the classifier/estimator-under-development and also to evaluate the design. Obviously, the training/evaluation data should be obtained from the process we are interested in.

In fact, all three subjects share the same key issue related to modelling, namely the selection of the appropriate generalization level. The empirical data are only an example of a set of possible measurements. If too much weight is given to the data at hand, the risk of overfitting occurs. The resulting model will depend too much on the accidental peculiarities (or noise) of the data. On the other hand, if too little weight is given, nothing will be learned and the model completely relies on the prior knowledge. The right balance between these opposite sides depends on the statistical significance of the data. Obviously, the size of the data is an important factor. However, the statistical significance also holds a relation with dimensionality.

Many of the mathematical techniques for modelling, tuning, training and evaluation can be shared between the three subjects. Estimation procedures used in classification can also be used in parameter estimation or state estimation, with just minor modifications. For instance, probability density estimation can be used for classification purposes and also for estimation. Data-fitting techniques are applied in both classification and estimation problems. Techniques for statistical inference can also be shared. Of course, there are also differences between the three subjects. For instance, the modelling of dynamic systems, usually called *system identification*, involves aspects that are typical for dynamic systems (i.e. determination of the order of the system, finding an appropriate functional structure of the model). However, when it finally comes to finding the right parameters of the dynamic model, the techniques from parameter estimation apply again.

Figure 1.4 shows an overview of the relations between the topics. Classification and parameter estimation share a common foundation indicated by 'Bayes'. In combination with models for dynamic systems (with random inputs), the techniques for
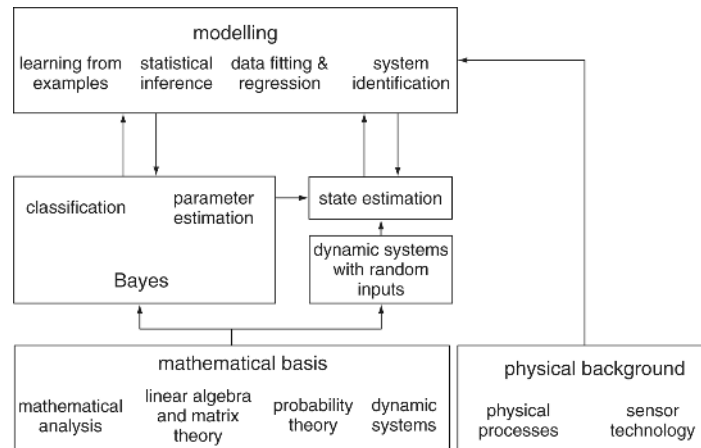
Figure 1.4 Relations between the subjects.

classification and parameter estimation find their application in processes that proceed in time, that is state estimation. All this is built on a mathematical basis with selected topics from mathematical analysis (dealing with abstract vector spaces, metric spaces and operators), linear algebra and probability theory. As such, classification and estimation are not tied to a specific application. The engineer, who is involved in a specific application, should add the individual characteristics of that application by means of the models and prior knowledge. Thus, apart from the ability to handle empirical data, the engineer must also have some knowledge of the physical background related to the application at hand and to the sensor technology being used.

All three subjects are mature research areas and many overview books have been written. Naturally, by combining the three subjects into one book, it cannot be avoided that some details are left out. However, the discussion above shows that the three subjects are close enough to justify one integrated book covering these areas.

The combination of the three topics into one book also introduces some additional challenges if only because of the differences in terminology used in the three fields. This is, for instance, reflected in the difference in the term used for 'measurements'.

In classification theory, the term 'features' is frequently used as a replacement for 'measurements'. The number of measurements is called the 'dimension', but in classification theory the term 'dimensionality' is often used.[1] The same remark holds true for notations. For instance, in classification theory the measurements are often denoted by **x**. In state estimation, two notations are in vogue: either **y** or **z** (MATLAB® uses **y**, but we chose **z**). In all cases we tried to be as consistent as possible.
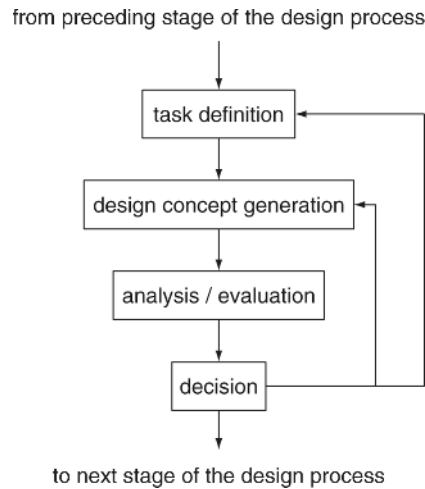
1.2 Engineering

The top-down design of an instrument always starts with some primary need. Before starting with the design, the engineer has only a global view of the system of interest. The actual need is known only at a high and abstract level. The design process then proceeds through a number of stages during which progressively more detailed knowledge becomes available and the system parts of the instrument are described at lower and more concrete levels. At each stage, the engineer has to make design decisions. Such decisions must be based on explicitly defined evaluation criteria. The procedure, the elementary design step, is shown in Figure 1.5. It is used iteratively at the different levels and for the different system parts.

An elementary design step typically consists of collecting and organizing knowledge about the design issue of that stage, followed by an explicit formulation of the involved task. The next step is to associate the design issue with an evaluation criterion. The criterion expresses the suitability of a design concept related to the given task, but also other aspects can be involved, such as cost of manufacturing, computational cost or throughput.

---

1 Our definition complies with the mathematical definition of 'dimension', i.e. the maximal number of independent vectors in a vector space. In MATLAB® the term 'dimension' refers to an index of a multidimensional array as in phrases like: 'the first dimension of a matrix is the row index' and 'the number of dimensions of a matrix is two'. The number of elements along a row is the 'row dimension' or 'row length'. In MATLAB® the term 'dimensionality' is the same as the 'number of dimensions'.

Figure 1.5 An elementary step in the design process (Finkelstein and Finkelstein, 1994).

from preceding stage of the design process

task definition

design concept generation

analysis / evaluation

decision

to next stage of the design process

Usually, there are a number of possible design concepts to select from. Each concept is subjected to an analysis and an evaluation, possibly based on some experimentation. Next, the engineer decides which design concept is most appropriate. If none of the possible concepts are acceptable, the designer steps back to an earlier stage to alter the selections that have been made there.

One of the first tasks of the engineer is to identify the actual need that the instrument must fulfil. The outcome of this design step is a description of the functionality, for example a list of preliminary specifications, operating characteristics, environmental conditions, wishes with respect to user interface and exterior design. The next steps deal with the principles and methods that are appropriate to fulfil the needs, that is the *internal functional structure* of the instrument. At this level, the system under design is broken down into a number of functional components. Each component is considered as a subsystem whose input/output relations are mathematically defined. Questions related to the actual construction, realization of the functions, housing, etc., are later concerns.

The functional structure of an instrument can be divided roughly into sensing, processing and outputting (displaying,

recording). This book focuses entirely on the design steps related to *processing.* It provides:

- Knowledge about various methods to fulfil the processing tasks of the instrument. This is needed in order to generate a number of different design concepts.
- Knowledge about how to evaluate the various methods. This is needed in order to select the best design concept.
- A tool for the experimental evaluation of the design concepts.

The book does not address the topic 'sensor technology'. For this, many good textbooks already exist, for instance see Regtien *et al.* (2004) and Brignell and White (1996). Nevertheless, the sensory system does have a large impact on the required processing. For our purpose, it suffices to consider the sensory subsystem at an abstract functional level such that it can be described by a mathematical model.

1.3 The Organization of the Book

Chapter 2 focuses on the introduction of PRTools designed by Robert P.W.Duin. PRTools is a pattern recognition toolbox for Matlab® freely available for non-commercial use. The pattern recognition routines and support functions offered by PRTools represent a basic set covering largely the area of statistical pattern recognition. In this book, except for additional notes, all examples are based on PRTools5.

The second part of the book, containing Chapters 3, 4 and 5, considers each of the three topics – classification, parameter estimation and state estimation – at a theoretical level. Assuming that appropriate models of the objects, physical process or events, and of the sensory system are available, these three tasks are well defined and can be discussed rigorously. This facilitates the development of a mathematical theory for these topics.

The third part of the book, Chapters 6 to 9, discusses all kinds of issues related to the deployment of the theory. As mentioned in Section 1.1, a key issue is modelling. Empirical data should be combined with prior knowledge about the physical process underlying the problem at hand, and about the sensory system used. For classification problems, the empirical data are often

represented by labelled training and evaluation sets, that is sets consisting of measurement vectors of objects together with the true classes to which these objects belong. Chapters 6 and 7 discuss several methods to deal with these sets. Some of these techniques – probability density estimation, statistical inference, data fitting – are also applicable to modelling in parameter estimation. Chapter 8 is devoted to unlabelled training sets. The purpose is to find structures underlying these sets that explain the data in a statistical sense. This is useful for both classification and parameter estimation problems. In the last chapter all the topics are applied in some fully worked out examples. Four appendices are added in order to refresh the required mathematical background knowledge.

The subtitle of the book, 'An Engineering Approach using MATLAB®', indicates that its focus is not just on the formal description of classification, parameter estimation and state estimation methods. It also aims to provide practical implementations of the given algorithms. These implementations are given in MATLAB®, which is a commercial software package for matrix manipulation. Over the past decade it has become the *de facto* standard for development and research in data-processing applications. MATLAB® combines an easy-to-learn user interface with a simple, yet powerful, language syntax and a wealth of functions organized in toolboxes. We use MATLAB® as a vehicle for experimentation, the purpose of which is to find out which method is the most appropriate for a given task. The final construction of the instrument can also be implemented by means of MATLAB®, but this is not strictly necessary. In the end, when it comes to realization, the engineer may decide to transform his or her design of the functional structure from MATLAB® to other platforms using, for instance, dedicated hardware, software in embedded systems or virtual instrumentation such as LabView.

MATLAB® itself has many standard functions that are useful for parameter estimation and state estimation problems. These functions are scattered over a number of toolboxes. The toolboxes are accompanied with a clear and crisp documentation, and for details of the functions we refer to that.

Most chapters are followed by a few exercises on the theory provided. However, we believe that only working with the actual algorithms will provide the reader with the necessary insight to

fully understand the matter. Therefore, a large number of small code examples are provided throughout the text. Furthermore, a number of data sets to experiment with are made available through the accompanying website.

1.4   Changes from First Edition

This edition attempts to put the book's emphasis more on image and video processing to cope with increasing interests on intelligent computer vision. More contents of most recent technological advancements are included. PRTools is updated to the newest version and all relevant examples are rewritten. Several practical systems are further implemented as showcase examples.

Chapter 1 is slightly modified to accommodate new changes in this Second Edition.

Chapter 2 is an expansion of Appendix E of the First Edition to accommodate the new changes of PRTools. Besides updating each subsection, the PRTools organization structure and implementation are also introduced.

Chapters 3 and 4 are, Chapters 2 and 3 in the First Edition, respectively.

Chapter 5 has now explicitly established the state space model and measurement model. A new example of motion tracking has been added. A new section on genetic station estimation has been written as Section 5.5. Further, an abbreviation of Chapter 8 of the First Edition has been formed as a new Section 5.6. The concept of 'continuous state variables' has been adjusted to 'infinite discrete-time state variables' and the concept of 'discrete state variables' to 'finite discrete-time state variables'. Several examples including 'special state space models' including "random constants', 'first-order autoregressive models', 'random walk' and 'second- order autoregressive models' have been removed.

In Chapter 6, Adaboost algorithm theory and its implementation with PRTools are added in Section 6.4 and convolutional neural networks (CNNs) are presented in Section 6.5.

In Chapter 7, several new methods of feature selection have been added in Section 7.2.3 to reflect the newest advancements on feature selection.

In Chapter 8, kernel principal component analysis is additionally described with several examples in Section 8.1.3.

In Chapter 9, three image recognition (objects recognition, shape recognition and face recognition) examples with PRTools routines are added.

1.5 References

Brignell, J. and White, N., *Intelligent Sensor Systems*, Revised edition, IOP Publishing, London, UK, 1996.

Finkelstein, L. and Finkelstein A.C.W., *Design Principles for Instrument Systems in Measurement and Instrumentation* (eds L. Finkelstein and K.T.V. Grattan), Pergamon Press, Oxford, UK, 1994.

Regtien, P.P.L., van der Heijden, F., Korsten, M.J. and Olthuis, W., *Measurement Science for Engineers*, Kogan Page Science, London, UK, 2004.